DOCUMENT RESUME

ED 288 843 SP 029 590

AUTHOR Nelson, David; Sandness, Wesley
TITLE Self-Evaluation in Student Teaching: A Reanalysis.
PUB DATE 19 Oct 86
NOTE 19p.; Paper presented at the Annual Convention of the Midwest Educational Research Association (Chicago, IL, October 19, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Age Differences; Cooperating Teachers; Higher Education; Preservice Teacher Education; *Self Evaluation (Individuals); *Sex Differences; *Student Teacher Attitudes; *Student Teacher Evaluation; Student Teacher Supervisors

ABSTRACT

This study examined how grade level and gender affected student teachers' self-ratings and faculty and field supervisors' ratings of student teachers' performance. The sample consisted of 47 preservice students in a medium-size Texas university. The Texas Performance Evaluation Instrument was administered in the second semester, at the end of a 12-week student teaching block. Independent ratings were obtained from 47 supervisors and 6 faculty supervisors immediately after the practice teaching session. Results indicated that, in general, both elementary and secondary school student teachers accorded themselves higher self-ratings than either field or college supervisors. Supervisors generally agreed in their ratings of secondary school teachers, but not elementary school teachers. Females' ratings tended to agree with supervisors' ratings more than males' ratings. (CB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED288843

Self-Evaluation in Student Teaching: A Reanalysis

A Paper Presented
at the

Annual Convention of the
Midwest Educational Research Association

Chicago, Illinois

October 19, 1986

David Nelson

Wesley Sandness

Pittsburg State University
Pittsburg, Kansas

SP 029 590

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Nelson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Self-Evaluation in Student Teaching: A Reanalysis

David Nelson
Wesley Sandness

During the past five years, eight states have adopted career ladder programs, nine states have endorsed mandated performance appraisal, and nine states have enacted teacher induction models. In such programs, the focus has shifted from superordinate evaluation to other types of evaluation, including student, peer, and self. Numerous books, articles, and research studies on the topic of self-evaluation (Bailey, 1981; Nelson and Blackburn, 1983; Olds, 1973; Pinola, 1982; Redfern, 1980; Richardson, 1975; Siedow, 1981) have encouraged practioners, so much so, that the practice is now being encouraged in many school districts. In other school systems, self-assessments have become integral components of performance objectives and/or clinical supervision models.

At the same time, self-evaluation has been recommended as part of a systems theory approach to supervision in colleges and departments of education (Bohning, 1978; Gitlin, 1981; Meyers, 1973; Sanders, 1974; Shapiro, 1977). Proposals of self-evaluation must be taken more seriously as school districts adopt programs of mandated performance appraisal. In fact, self-evaluation skills could be easily introduced in courses related to professionalism (methods), and later implemented in various levels of field-related experiences such as student teaching and possibly teacher induction. Skeptics, on the other hand, argue against the concept on the tenuous grounds that student teachers are too inexperienced or too immature to make realistic self-evaluations.

The issues are further confused by a paucity of empirical investigations in the field which have rendered conflicting results. On the one hand, Chiu (1975) and Pease (1975) found that student teachers did not, as a rule, have distorted perceptions of their teaching abilities. In other words, student self-ratings and supervisory ratings were strongly correlated. On the other hand, Nelson and Blackburn (1983) found that student teachers continually overrated their performance in relation to both college faculty and field supervisors.

Differences in research design, sampling procedures, and instrumentation may have accounted for some of the discrepancies. In the earlier study, Nelson and Blackburn (1983) failed to consider two crucial factors. the first is grade level (elementary versus secondary student teachers) and the second is gender (female versus male students). Both Chiu (1975) and Pease (1975) sampled only elementary candidates, and most respondents were female. In contrast, Nelson and Blackburn (1983) maintained a balanced student teacher representation in terms of both grade level and gender. This distinction raises a significant question: Could so-called differences in rater groups be attributed to the unequal distribution of student teachers' characteristics in the studies cited?

The purpose of this investigation is to examine this question, first as it relates to elementary versus secondary student teachers, and second as it relates to female versus male student teachers. Specifically, first by grade level and second by gender, the research questions become: (1) Do elementary and secondary student teacher self-ratings generally agree with supervisors ratings? (2) Do elementary or secondary student teacher self-ratings agree more closely

with ratings made by college faculty or by field supervisors? (3) Do faculty supervisors assess elementary and secondary student teaching performance in the same way as field supervisors? (4) Do female or male student teacher self-ratings agree more closely with ratings made by academic or by field supervisors? (5) Do faculty supervisors assess female and male student teacher performance in the same way as field supervisors?

Methodology

Data Collection

The sample consisted of 47 preservice students in a medium size Texas university. Two student teaching sections, one elementary (23 students) and one secondary, (24 students) were chosen to participate in the study. Confounding effects were presumably minimized by maintaining a balanced distribution of female (25) and male (22) candidates within the two sections. Both groups were staffed by four faculty members and two graduate students who stayed with the groups throughout a series of carefully organized activities (competencies), including student teaching.

The Texas Performance Evaluation Instrument (Heeney and Nelson, 1983) was administered to all students in the second semester, 1986, at the end of a twelve-week student teaching block. All students by this time had completed eight weeks of student teaching experience. Independent ratings on student teachers were obtained from respective field supervisors (N=47) and faculty supervisors (N=6) immediately after the same practice teaching session. Instructions to the three rater

groups explained reasons for the study and carefully stressed that this evaluation was entirely separate from, and independent of, the normal assessment procedures employed by the institution. Anonymity was guaranteed to all participants prior to administration of the Instrument.

Instrumentation

The Texas Performance Evaluation Instrument consists of 78 statements which are clustered into seven categories: Instruction (18 items), Student Relations (18 items), Learning Climate (6 items), Curriculum (8 items), Relations to Peers and Administration (8 items), Professional Qualities (12 items), Personal Qualities (8 items). Student teachers were asked to rate their own performance and field supervisors and faculty supervisors were asked, in turn, to rate the performance of each candidate on a six-point scale ranging from: (1) strongly disagree to (6) strongly agree. A no-data option was provided for raters who lacked sufficient information to assess performance on a specific item. Sample items included:

(1) Instruction:

1. Plans and organizes for effective teaching.

12. Adjusts the "difficulty of instruction" to the ability of the students.

(2) Student Relations:

27. Is fair, consistent, and objective in dealing with students.

33. Does not have discipline problems.

(3) Learning Climate:

38. Provides an appropriate climate in which students can learn and interact.

40. Provides a physical environment which is conducive to learning.

(4) Curriculum:

45. Selects and organizes instructional activities appropriate to course objectives.

50. Evaluates the instructional outcomes of a course by means of a variety of appropriate measures.

(5) Relations with Peers and Administration:

54. Cooperates with fellow teachers in the development of effective teaching practices.

58. Supports administrative efforts to improve instruction.

(6) Professional Qualities:

62. Is open to constructive criticism.

69. Maintains an appropriate level of self-control and emotional stability.

(7) Personal Qualities:

72. Is enthusiastic.

78. Serves as a model for students.

Data concerning the reliability and validity of the instrument have proven it to be consistent over time and accurate in measuring what it purports to measure. For example, Cronback alpha reliability coefficients, on the present sample, range from a low of .71 (Instruction) to a high of .95 (Relations to Peers and Administration) (Heeney and Nelson, 1983). As well, the instrument has been shown to possess adequate convergent and content validity (Heeney and Nelson, 1983).

Data Analysis

Several analyses were conducted to fulfill the objectives of the study. First, mean scores were calculated for each category within the Instrument and for overall performance. Second, t-tests were performed to determine specific differences between the three pairs of rater groups. Third, Pearson product-moment correlation coefficients were computed to assess the degree of congruence between rater groups for male and female teachers.

Results

The first research question examines the degree of consistency between elementary and secondary student teacher self-ratings and the two sets of supervisory ratings. Figures 1 and 2 illustrate the findings for both the elementary and secondary grade levels respectively. Results show that, with one exception, both elementary and secondary preservice teachers accorded themselves higher mean self-ratings than either the field supervisors or the college supervisor. The one

FIGURE 1.

MEANS, STANDARD DEVIATIONS, AND T-TESTS FOR ELEMENTARY, SELF, SUPERVISOR, AND FACULTY RATINGS OF STUDENT TEACHING PERFORMANCE (N=23)

Category	Self		Supervisor		Faculty		t-values		
	M	SD	M	SD	M	SD	Self/ Supv.	Self/ Faculty	Supv/ Faculty
(1) Instruction	5.36	.40	4.93	.82	5.05	.59	2.83*	2.69*	.90
(2) Student Relations	5.48	.34	5.10	.67	5.01	.49	2.79*	4.73***	.76
(3) Learning Climate	5.53	.36	5.01	.89	4.68	.76	3.00*	4.91***	1.47
(4) Curriculum	5.51	.39	4.86	.84	5.00	.58	3.89***	4.06***	1.04
(5) Relations to Peers and Admin.	5.60	.77	5.64	.74	5.19	.77	.24	2.46*	5.00***
(6) Professional Qualities	5.77	.23	5.55	.58	5.10	.76	1.98	4.62***	3.74***
(7) Personal Qualities	5.83	.24	5.61	.70	5.00	.71	1.51	5.79***	4.82***
Overall Performance	5.58	.27	5.24	.66	5.02	.60	2.70	5.30***	2.17*

*p < .05
**p < .01
***p < .001

FIGURE 2.

MEANS, STANDARD DEVIATIONS, AND T-TESTS FOR SECONDARY SELF, SUPERVISOR, AND FACULTY RATINGS OF STUDENT TEACHING PERFORMANCE (N=24)

Category	Self M	Self SD	Supervisor M	Supervisor SD	Faculty M	Faculty SD	t-values Self/ Supv.	t-values Self/ Faculty	t-values Supv/ Faculty
(1) Instruction	5.56	.47	4.64	.97	4.94	.80	4.36***	3.61***	1.80
(2) Student Relations	5.72	.34	4.71	1.03	5.04	.65	4.52***	5.11***	1.87
(3) Learning Climate	5.65	.56	4.82	.92	4.94	.87	4.64***	3.67***	.65
(4) Curriculum	5.68	.41	4.60	.96	5.15	.94	5.35***	2.68*	3.53***
(5) Relations to Peers and Admin.	5.88	.21	5.33	1.23	5.19	1.15	2.32*	3.33**	.81
(6) Professional Qualities	5.85	.19	5.07	1.12	4.98	.89	3.35***	5.03***	.54
(7) Personal Qualities	5.75	.44	5.10	1.16	4.93	.93	2.57*	4.40***	.97
Overall Performance	5.73	.34	4.98	.94	5.02	.81	4.13***	4.35***	1.07

*p < .05
**p < .01
***p < .001

exception is between self and associates on Item 5 (Relations to Peers and Administrators) at the elementary level. Obviously, there are differences in perception of abilities between student teachers, on the one hand, and both sets of supervisors on the other.

The second research question examines these differences as they relate to the two types of supervisors--field and college. Figure 1 reveals that at the elementary level there is more disagreement between self and faculty than between self and field supervisors. Indeed, differences between self and faculty assessments are significant across all categories and reveal unanimous disagreement in overall performance ($p<.001$). However, ratings between self and field supervisors showed significant differences for only four categories and for overall performance ($p<.05$). This finding leaves three areas of agreement: Relations to Peers and Administration, Professional Qualities, and Personal Qualities. The secondary preservice teachers (Figure 2), results are even more disparate. Throughout, there is total disagreement between the student teachers and the two sets of supervisors.

The third research question examines the degree of congruity between faculty and field supervisory ratings. At the elementary level, Figure 1 indicates agreement on four factors (Instruction, Student Relations, Learning Climate, and Curriculum) and disagreement on three items (Relations to Peers and Administration, Professional Qualities, and Personal Qualities) and on overall performance ($p<05$). In contrast, with the exception of Curriculum, there is consensus on all factors at the secondary level. This finding seems to suggest that faculty and associates at the secondary level are far more congruent in

their perceptions of student teacher abilities than are their counterparts at the elementary level.

The fourth research question considers male and female self-ratings as they relate to academic and field supervisor ratings. In Figure 3, Pearson product-moment correlation coefficients between rater groups show little agreement between male candidates and the two groups of supervisors. Only one coefficient (Learning Climate) is significant for self/supervisor ratings ($r=.49$, $p<.05$) and none are significant for the self/faculty ratings. Yet, female student teachers agree with faculty supervisors on four factors (Curriculum, Relations to Peers and Administration, Professional Qualities, and Personal Qualities) and with field supervisors on three factors (Instruction, Curriculum, Relations to Peers and Administration). Moreover, the overall performance scores for female candidates are significantly correlated for self/supervisor ($r=.39$, $p<.05$) and self/faculty ratings ($r=.48$, $p<.01$), while the correlations for male student teachers are viewed more favorably than male student teachers by both sets of supervisors. In particular, they are seen as better curriculum developers and as being more able to relate to peers and principals.

The last research question deals with consistency between faculty and supervisor ratings for female and male candidates. As shown in Figure 3, all of the correlations are significant ($p<.01$), thus indicating a high degree of supervisory consensus on the teaching abilities of both male and female participants.

FIGURE 3.

PEARSON PRODUCT-MOMENT CORRELATION COEFFICIENTS
BETWEEN SELF, SUPERVISOR, AND FACULTY
RATINGS OF MALE AND FEMALE
STUDENT TEACHING PERFORMANCE

Category	MALE (N=22)			FEMALE (N=25)		
	Self/ Supvr.	Self/ Faculty	Supvr/ Faculty	Self/ Supvr.	Self/ Faculty	Supvr/ Faculty
(1) Instruction	.01	.09	.59**	.34*	.32	.60***
(2) Student Relations	.01	.25	.59**	.01	.29	.59***
(3) Learning Climate	.49*	.19	.53**	.23	.19	.63***
(4) Curriculum	.16	.09	.61***	.40*	.42*	.54**
(5) Relations to Peers and Admin.	.01	.33	.82***	.74***	.51**	.70***
(6) Professional Qualities	.04	.13	.75***	.22	.48**	.60***
(7) Personal Qualities	.05	.20	.75***	.29	.34*	.61***
Overall Peformance	.07	.21	.79***	.39	.48**	.71***

*p < .05
**p < .01
***p < .001

Discussion

In the initial study, Nelson and Blackburn (1983) observed that "most correlation coefficients between self-ratings and supervisor ratings are low and not significant." Division of students by grade level did not alter this finding. Accordingly, at the elementary level, there appears to be significant agreement on the self/faculty ratings, but some disagreement on the self/supervisor ratings. At the secondary level, differences are even more acute as students and supervisors disagree on all factors. In short, there is little agreement between either the elementary or secondary student teachers and their respective supervisors but a considerable amount of agreement between the supervisors themselves.

Undoubtedly, the most interesting results of this inquiry center around gender differences in terms of rater performance. Female student teachers were continually viewed more favorably than male student teachers by both supervisors and faculty. This means, in essence, that when the sample is restricted to only female students. there is substantial agreement among the rater groups (self/associate/faculty). Thus, when we confine the sample to only female students, there is substantial agreement among the rater groups (self/supervisor/faculty). Thus, when we confine the sample to predominantly female students, the present results are in support of Chiu (1975) and Pease (1975), not Nelson and Blackburn (1983). Please recall, though the first two studies limited their samples to predominantly female respondents, the latter did not.

The intriguing question is, of course, why the sexual disparity in supervisory ratings? Could it be that male students, as a group, overestimate their teaching abilities? A t-test between mean ratings for male and female students across categories and for overall performance negated such a claim. In fact, there were no significant differences between mean ratings for the two sexes on any one category or on overall performance. Translated, both male and female respondents rated themselves about equal in terms of teaching ability. Consequently it would appear that male student teachers are not overstating their competencies if indeed they are as proficient as their female rivals. Assuming this to be the case, then, their supervisors are clearly underestimating their abilities. Once again, the question becomes, why? Why do academic and field supervisors continually rate female preservice teachers higher than male teacher education candidates?

Well, first let us look at the female student teacher. She has the advantage of compatible sex-matching which, in turn, results in much positive feedback (external reinforcement builds confidence), she identifies closely with the teaching profession (feminine or masculine task), and externalizes a high degree of success which is reflected in her self-rating. The male student teacher, on the other hand, receives much negative feedback due to incompatible sex-matching. Nevertheless, his cognitive set, sex-matching patterns of rater to ratee, tells him to ignore such feedback as it is an external locus of control. Indeed, he has been conditioned to automatically attribute criticism to lack of effort or rater bias, not ability. Moreover, his convictions are strengthened through professional socialization which

fosters male superiority, and consequently a high level of self-esteem. Obviously, then, the sex of the student teacher is a key factor in a study of this sort and, as such, should be included in future research studies.

Conclusions

The present investigation may have some profound implications for the effective supervision of student teaching. First, a balancing in the sex-matching patterns of rater to ratee might help to facilitate more impartial supervision. This would seem to be particularly critical in an institution where the clientele is predominantly female and the faculty predominantly male. Without equitable balancing, male candidates could be at an extreme disadvantage. Second, it would appear that the amount and type of feedback provided to student teachers should be carefully monitored. For those who wish to increase performance, it should be clear that negative feedback is not usually effective. In fact, it can be dramatically counter-productive, especially when the recipient is programmed to internalize failure (Layden & Ickes, 1978). Does this mean that disapproval should be objective and directed to the task at hand. It should be clearly stated, precise. constructive and, perhaps most importantly, given in a nonthreatening environment. Finally, the authors propose that supervisors should not deliberately challenge or discourage students who have inflated self-concepts; provided, of course, they are within the norm. Certainly, evidence suggests that individuals with positive self-images perform better, on the average, than those with negative self-images. In particular, those

with a high level of self-worth are, as a rule, more easily motivated, better adjusted, take greater pride in their work, and are healthier from both a physical as well as emotional perspective (Korman, 1977; Zander, 1963).

Bibliography

Bailey, D.G. (1981). Learning how to self-critique using audiotape and videotape in teacher self-assessment. Educational Technology, 21 (2), 41-45.

Bohning, G. (1978). Subjective judgment pitfalls in evaluating student teachers. Teacher Educator, 14 (1), 13-15.

Brannigan, G.G. and Tolor, A. (1971). Sex differences in adaptive styles. Journal of Genetic Psychology, 119, 143-149.

Brophy, B. (1976). Teacher evaluation: Ultimately self-evaluation. Comment on Education, 6 (4), 6-9.

Chiu, L.H. (1975). Influences of student teaching on perceived teaching competence. Perceptual and Motor Skills, 40, 872-874.

Dweck, C.S. and Bush, E.S. (1976). Sex differences in learned helplessness: I. Differential debilitation with peer and adult evaluators. Developmental Psychology, 12, 147-156.

Gitlin, A. (1981). Horizontal evaluation: An approach to student teaching supervision. Journal of Teacher Education, 32 (5), 47-50.

Heeney, W. and Nelson, D. (1983). Assessing the Texas performance instrument. (ERIC Document Reproduction Service No. ED 122 489)

Jacobson, M. and Effertz, J. (1974). Sex roles and leadership perceptions of the leaders and the led. Organizational Behavior and Human Performance, 12, 383-396.

Korman, A.K. (1977). Organizational Behavior. Englewood Cliffs, N.J.: Prentice-Hall.

Layden, M.A. and Ickes, W.J. (1978). Self-esteem and sex differences in attributional style: Effect upon performance. (ERIC Document Reproduction Service No. ED 169 406)

Lewis, F. (1978). The Lewis self-evaluation scale. Phi Delta Kappan, 59 (10), 686-690.

Meyers, P.E. (1973). The real crux of supervision. Contemporary Education, 44 (3), 140-144.

Nelson, D. and Blackburn, K. (1983). Self-Assessments of Student Teachers. Paper presented at the American Educational Research Association, Chicago.

Olds, R. (1973). Self-evaluation for teachers and administrators. Worthington, Ohio: School Management Institute.

Pease, D. (1975) . Comparing faculty and school supervisor (1), 91-94.

Pinola, M. (1982). Are you a teacher? A short quiz for self-evaluation. Education, 102, 384-388.

Redfern, G.B. (1980). Evaluating teachers and administrators: A performance objectives approach. Boulder, CO: Westview Press.

Richardson, E.T. (1975). A self-evaluation instrument. NASSP Bulletin, 59, 7-17.

Sanders, J.T. (1974). Toward a rationale for student teaching. Teacher Education, 7, 18-24.

Shapiro, P.P. (1977). The evaluation of student teachers. Teacher Education, 11, 40-48.

Siedow, M.D. (1981). Teacher self-evaluation for instructional improvement. Clearing House, 55 (4), 176-179.

Sullivan, J.P. (1978). Evaluating teacher performance. Arlington, VA: Educational Research Service

Windom, C.S. and Burke, B.W. (1978). Performance, attitudes and professional socialization of women in academia. Sex Roles, 4 (4), 549-562.

Zander, A.F. (1963). Research on self-evaluation, feedback and threats to self-esteem. In A. Zander (ed.), Performance appraisals: Effects on employees and their performance (pp.5-17). Ann Arbor, MI: The Foundation for Research in Human Behavior.